Truncated distributions

---

35 marks

The questions here are designed to explore some basic characteristics of, and differences between, probability distributions and the random realizations from them. See `help("Distributions")` for those built into `R` .

1. Suppose we have a continuous random variable X with distribution function $F_X(x) = Pr(X \le x)$ and quantile function $Q_X(p) = F_X^{-1}(p)$. That is $p = F_X(x) = Pr(X \le x)$ and $p = Pr(X \le Q_X(p)) = F_X(Q_X(p)) = F_X(F_X^{-1}(p)) = p$.

   We can define a random variable Y having cumulative distribution function

$$G_Y(y) = \begin{cases} 0 & y < a \\ \frac{F_X(y) - F_X(a)}{F_X(b) - F_X(a)} & a \le y \le b \\ 1 & y > b \end{cases}$$

   where $-\infty \le a < b \le \infty$ and X is a continuous random variable as above.

   That is, Y has the same distribution as X **except** that it is **truncated** on the left at a and on the right at b. Unlike X, Y cannot take values less than a or larger than b.

   a. *(3 marks)* Mathematically show that the random variable W defined

$$W = Q_X(F_X(a) + U \times (F_X(b) - F_X(a)))$$

   where U is a uniform random variable $U \sim U(0, 1)$ has the same distribution as Y.

   We can apply the $F_X()$ to W and disscuss the three cases. Note that $U \in [0, 1]$.

   Case 1: $w < a$

$$Pr(F_X(a) + U \times (F_X(b) - F_X(a)) \le F_X(w)) = Pr(U \times (F_X(b) - F_X(a)) \le F_X(w) - F_X(a)) = 0$$

   Case 2: $a \le w \le b$

$$Pr(F_X(a) + U \times (F_X(b) - F_X(a)) \le F_X(w)) = Pr(U \le \frac{F_X(y) - F_X(a)}{F_X(b) - F_X(a)})$$

   Case 3: $b < w$

$$Pr(F_X(a) + U \times (F_X(b) - F_X(a)) \le F_X(w)) = Pr(U \times (F_X(b) - F_X(a)) \le F_X(w) - F_X(a)) = 1$$

   b. *(15 marks)* Here you are to write a function `truncate()` of the form

```r
truncate <- function(ddist = dnorm, pdist = pnorm, qdist = qnorm, a = -Inf, b = Inf) {
  list(
    # density
    ddist = function(x) {
      den <- c()
      for (i in 1 : length(x)){
```

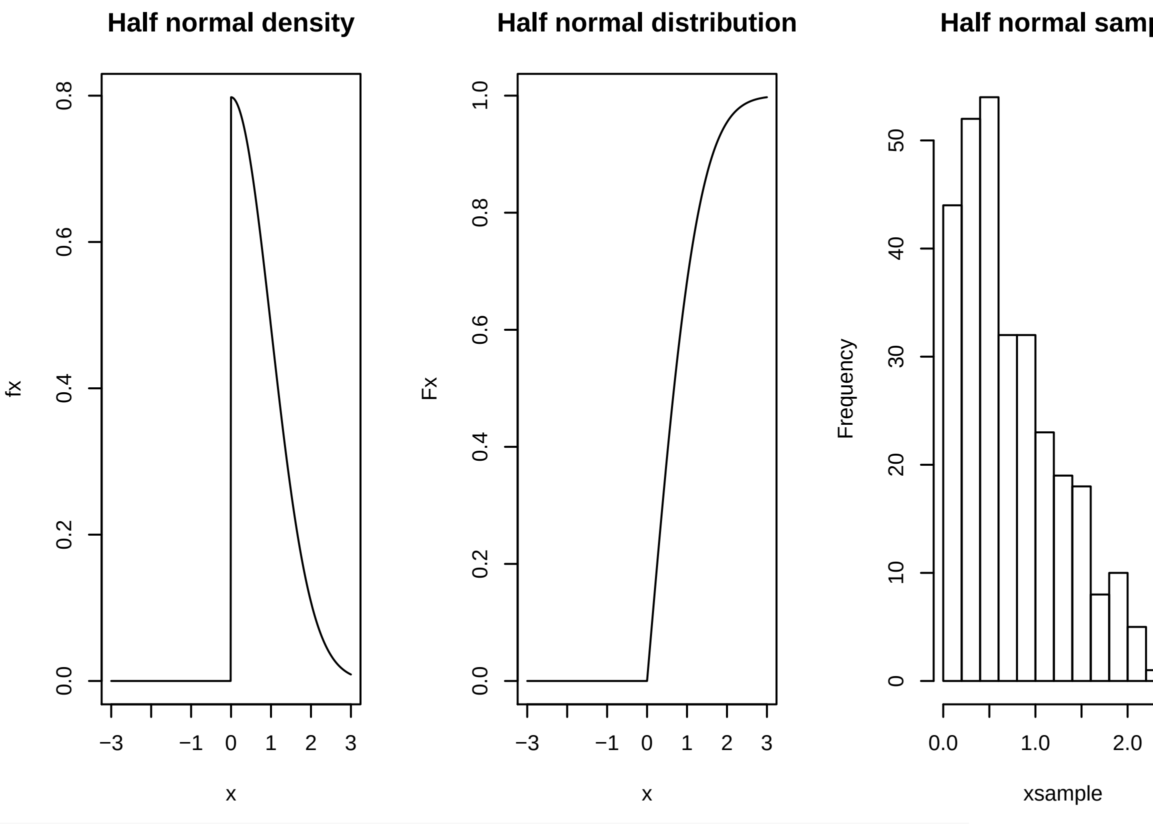
```r
        if (x[i] >= a & x[i] <= b){
          den[i] = (ddist(x[i]) / (pdist(b) - pdist(a)))
        } else {
          den[i] = 0
        }
      }
      return(den)
    },
    # distribution
    pdist = function(x, mean = 0,sd = 1) {
      distribution <- c()
      for (i in 1:length(x)){
        if (x[i] < a){
          distribution[i] = 0
        } else if (x[i] > b) {
          distribution[i] = 1
        }
        else {
          # calc by pdist(x)-pdist(a)]/[pdist(b)-pdist(a)]
          distribution[i] = (pdist(x[i],mean = mean,sd = sd) - pdist(a,mean = mean,sd = sd)) / (pdist
        }
      }
      return(distribution)
    },
    rdist = function(n,mean = 0, sd = 1) {
      return(qdist(pdist(a,mean = mean,sd = sd) + runif(n)*(pdist(b,mean = mean,sd = sd) - pdist(a,mea
    }
  )
#list(ddist, pdist, rdist)
}
```

```
where `ddist`, `pdist`, and `qdist` refer to functions which  calculate the density $f_X(x)$, dist

The function `truncate()` is to return a list with components named `ddist`, `pdist` and `rdist` c

That is, the following should work for the half-normal distribution.
```

```r
    half_normal <- truncate(a = 0)
    xsample <- half_normal$rdist(300)
    x <- seq(-3, 3, 0.01)
    fx <- half_normal$ddist(x)
    Fx <- half_normal$pdist(x)
    oldPar <- par(mfrow = c(1,3))
    plot(x, fx, type = "l", main = "Half normal density")
    plot(x, Fx, type = "l", main = "Half normal distribution")
    hist(xsample, main = "Half normal sample")
```



results-1.pdf results-1.bb

```
par(oldPar)
```

Hand in the above plots with your code.

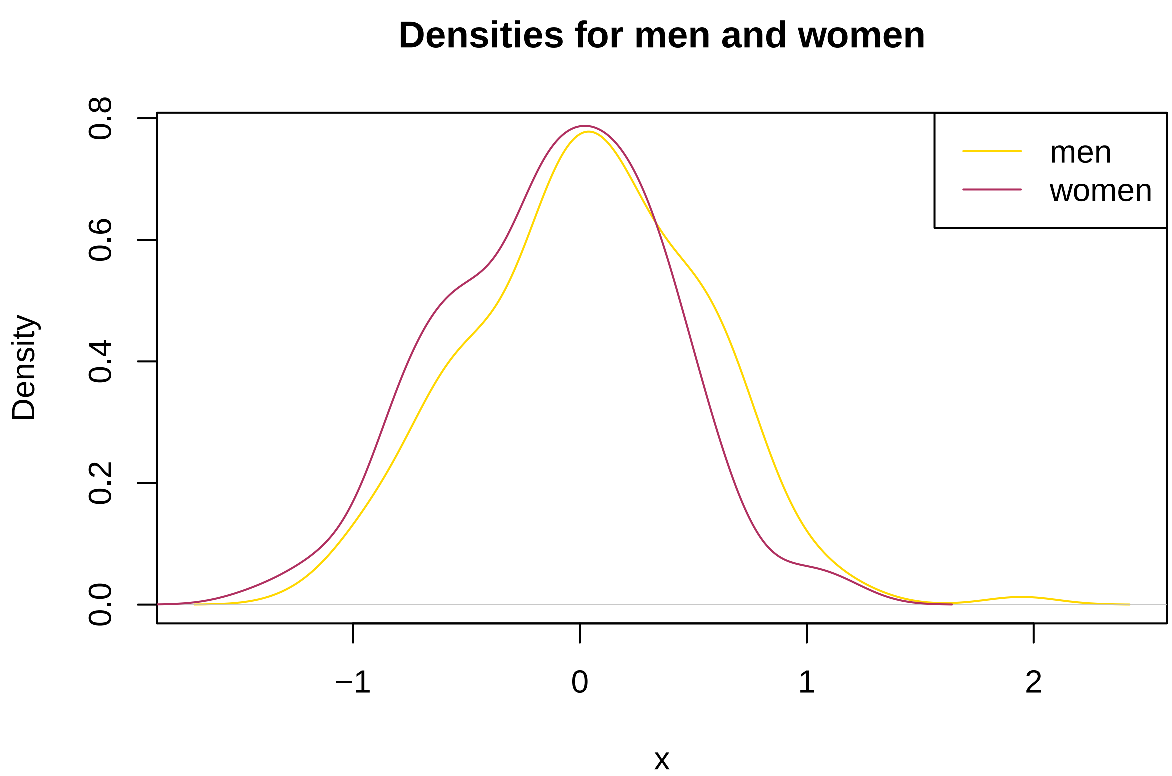
c. A 2011 article by Gil Greengross and Geoffrey Miller of the University of New Mexico was entitled "

They tested the sense of humour of 400 university students (200 men, 200 women) using a standardiz

The distributions of humour between men and women seems to be significantly different but what doe

i. *(3 marks)* On a single (nicely labelled with a legend) draw the densities (in different colour

```
men <- rnorm(200, mean = 0.09, sd = 0.49)
women <- rnorm(200, mean = -0.09, sd = 0.49)
density_men = density(men)
density_women = density(women)
plot(density_men, col = "gold", xlab = 'x', main = "Densities for men and women")
lines(density_women, col = "maroon")
legend("topright", legend = c("men", "women"), col = c("gold", "maroon"), lwd = c(1,1))
```




```
ii. *(4 marks)* Generate a random sample of 1000 scores from each of these distributions and save t

    Based on your sample, estimate the following

    - the average humour ability of the men
    - the average humour ability of the women
    - the probability that the man will be funnier than the woman (at least as measured by this sca
```

```r
x = rnorm(1000, mean = -0.09, sd = 0.49)
y = rnorm(1000, mean = 0.09, sd =0.49)
results <- data.frame(women = x, men = y)

print("- the average humour ability of the men")
```

```
## [1] "- the average humour ability of the men"
```

```r
print(mean(men))
```

```
## [1] 0.0387097
```

```r
print("- the average humour ability of the women")
```

```
## [1] "- the average humour ability of the women"
```

```r
print(mean(women))
```

```
## [1] -0.1060419
```

```r
print('- the probability that the man will be funnier than the woman (at least as measured by this scale
```

```
## [1] "- the probability that the man will be funnier than the woman (at least as measured by this scal
```

```r
# count one by one....
counter = 0
```

```r
for(i in 1:1000){
  if(y[i] > x[i]){
  counter <- counter + 1
  }
}
prob <- counter / 1000
print(prob)
```

```
## [1] 0.61
```

```
     iii. *(4 marks)*  [Are women funny?](https://www.theguardian.com/commentisfree/2014/mar/04/are-wome
          Generate 1000 pseudo random scores `y` from the truncated distribution for men and another 10(

            - the average humour ability of the men
            - the average humour ability of the women
            - the probability that the man will be funnier than the woman (at least as measured by thi:
```

```r
      funny <- truncate(a = 1.07)
      men <- funny$rdist(1000, mean = 0.09, sd = 0.49) #men
      women <- funny$rdist(1000, mean = -0.09,sd = 0.49) #women=
      res <- data.frame(women = women, men = men)
      mean_man <- mean(men)
      mean_women <- mean(women)
      count <- 0
      for(i in 1:1000){
        if(men[i] > women[i]){
        count <- count + 1
        }
      }
      prob <- count/1000
      print('- the average humour ability of the men')
```

```
## [1] "- the average humour ability of the men"
```

```r
      print(mean_man)
```

```
## [1] 1.254786
```

```r
      print('- the average humour ability of the women')
```

```
## [1] "- the average humour ability of the women"
```

```r
      print(mean_women)
```

```
## [1] 1.232599
```

```r
      print('- the probability that the man will be funnier than the woman (at least as measured by thi:
```

```
## [1] "- the probability that the man will be funnier than the woman (at least as measured by this scal
```

```r
      print(prob)
```

```
## [1] 0.528
```

```
    iv. *(2 marks)* What conclusions do you draw about the differences between the humour of men and wo

      There is no significant difference between the humor of men and women.
```

```
v. *(4 marks)* Repeat part iii, again conditioning on considering only individuals with a "humour al
```

```r
funny2 <- truncate(a = 1.07)
  men <- funny2$rdist(1000,mean = 0.09,sd = 0.49 * 1.1) #men
  women <- funny2$rdist(1000,mean = 0.09,sd = 0.49) #women

  result <- data.frame(women = men, men = men)

  mean_men <- mean(men)
  mean_women <- mean(women)
  count <- 0
  count <- 0
  for(i in 1:1000){
    if(men[i] > women[i]){
    count <- count + 1
    }
  }
  prob <- count/1000
  print('- the average humour ability of the men')
```

```
## [1] "- the average humour ability of the men"
```

```r
  print(mean_man)
```

```
## [1] 1.254786
```

```r
  print('- the average humour ability of the women')
```

```
## [1] "- the average humour ability of the women"
```

```r
  print(mean_women)
```

```
## [1] 1.253287
```

```r
  print('- the probability that the man will be funnier than the woman (at least as measured by thi
```

```
## [1] "- the probability that the man will be funnier than the woman (at least as measured by this sca
```

```r
  print(prob)
```

```
## [1] 0.56
```